

Unfamiliar face matching with frontal and profile views

Robin SS Kramer

School of Psychology, University of Lincoln, UK; Department of Psychology, Trent University, Canada; Department of Psychology, University of York, UK

Michael G Reynolds

Department of Psychology, Trent University, Canada

Corresponding author:

Robin Kramer, School of Psychology, University of Lincoln, Lincoln, LN6 7TS, UK.

Email: remarknibor@gmail.com

Abstract

Research has systematically examined how laboratory participants and real-world practitioners decide whether two face photographs show the same person or not, using frontal images. In contrast, research has not examined face matching using profile images. In Experiment 1, we ask whether matching unfamiliar faces is easier with frontal compared with profile views. Participants completed the original, frontal version of the Glasgow Face Matching Test, and also an adapted version where all face pairs were presented in profile. There was no difference in performance across the two tasks, suggesting that both views were similarly useful for face matching. Experiments 2 and 3 examined whether matching unfamiliar faces is improved when both frontal and profile views are provided. We compared face matching accuracy when both a frontal and a profile image of each face was presented, with accuracy using each view alone. Surprisingly, we found no benefit when both views were presented together in either experiment. Overall, these results suggest that either frontal and profile views provide substantially overlapping information regarding identity or participants are unable to utilise both sources of information when making decisions. Each of these conclusions has important implications for face matching research and real-world identification development.

Keywords

unfamiliar faces, face matching, frontal view, profile view, individual differences

Introduction

Numerous studies have established that unfamiliar face matching, where viewers must decide whether images depict the same person or different people, is both difficult and error-prone (Bruce, Henderson, Newman, & Burton, 2001; Bruce et al., 1999; Megreya & Burton, 2006, 2008). Depending on the task, error rates typically range from 10% to 30%. Researchers have shown that decisions involving unfamiliar faces are closely bound to the visual properties of the particular images (Hancock, Bruce, & Burton, 2000) and, as such, the process is qualitatively different from familiar face matching (Megreya & Burton, 2006).

Given that matching images of unfamiliar faces is limited to information in the pictures themselves, how might performance be improved on this task? Although research into this question is ongoing, few solutions have been found so far. In terms of the process, we know that working together in pairs (Dowsett & Burton, 2015) or aggregating the responses of groups of individuals (White, Burton, Kemp, & Jenkins, 2013) can increase accuracy. Regarding the materials, evidence suggests that using computer-generated averages or arrays of instances can improve performance over comparison with a single image (White, Burton, Jenkins, & Kemp, 2014). In addition, cropping images so that only the internal features are visible can increase accuracy for

difficult comparisons (Kemp, Caon, Howard, & Brooks, 2016). Importantly, all these studies have involved comparisons between frontal images only.

Face matching with profile views

Little is known about unfamiliar face matching with non-frontal images. On the one hand, this is not surprising, given that all forms of photo identification that the authors are aware of (e.g., passports, driving licences) utilise images taken front-on, which seems to imply a widely held belief that the frontal view is superior and that the profile view does not add sufficient new information to warrant inclusion. Furthermore, researchers have yet to investigate how accurately people are able to match faces shown in profile views, perhaps implying that the research community may also believe that profile views are unimportant. On the other hand, the absence of research examining how profile views are matched is surprising, given that mugshots were originally standardised by Alphonse Bertillon in the 1880s so as to contain both a frontal and a profile view. This standard was rapidly adopted across a number of countries including the USA, suggesting a widespread belief (at least in the justice system) that both views were important for identification. Despite this format for arrest photographs, forms of identification commonly used around the world include only frontal views.

The purpose of the present paper is to provide some initial insight into how face matching occurs with profile views. To this end, two questions are addressed. In

Experiment 1, we ask whether matching unfamiliar faces is easier with frontal compared with profile views. In Experiments 2 and 3, we ask whether matching unfamiliar faces is improved when both frontal and profile views are provided.

Experiment 1

Is unfamiliar face matching easier for frontal in comparison with profile views? As noted above, the widespread inclusion of frontal, but not profile view, photos in identification documents suggests that face matching is generally assumed to be better for frontal compared with profile views of unfamiliar faces. However, there is some evidence from studies of face recognition (which differs from matching in that memory plays a role) that when the same view is presented at study and test, there is no difference in performance for frontal pairs and profile image pairs (Liu & Chaudhuri, 2002; Logie, Baddeley, & Woodhead, 1987). This suggests that both frontal and profile views may contain similar amounts of information for supporting face matching.

In order to assess whether there is a performance advantage for the frontal view during face matching, participants in Experiment 1 completed two versions of the Glasgow Face Matching Test (GFMT; Burton, White, & McNeill, 2010). In one block of trials, they completed the standard frontal view version of the test. In another block of trials, they completed a new profile view version of the GFMT. Critically, in order to make it possible to directly compare performance in the two blocks, the frontal- and

profile-view versions of the GFMT were matched by using the same identities, cameras and camera distances, clothing, and hair-styles.

Methods

Participants

Fifty-four students (49 women; age $M = 19.22$ years, $SD = 2.12$; 72.22% self-reported ethnicity as White) at Trent University took part in exchange for course credits. All participants in Experiments 1 and 2 provided written informed consent and were verbally debriefed at the end of the experiment. Sample size was based on past research using a face matching paradigm (Dowsett & Burton, 2015; Estudillo & Bindemann, 2014), although we continued to run participants until the end of a pre-determined data collection run. Trent University's ethics committee approved all experiments presented here, which were carried out in accordance with the provisions of the World Medical Association Declaration of Helsinki.

Stimuli

Images from the short version of the GFMT (Burton et al., 2010) were used as frontal view stimuli. These comprised forty pairs of male and female faces viewed front on, where half the pairs were 'match' trials (different images of the same person) and half were 'mismatch' trials (different people with a similar appearance). To create 'match'

trials, two different images were collected for each face – a high-quality digital photograph and a still frame from a high-quality video. In both cases, camera distance (90 cm) and expression (neutral) were unchanged, and so only the camera itself differed. The images were taken approximately 15 minutes apart. Presenting two similar, but different, images of a face prevents participants from performing simple picture matching (Jenkins & Burton, 2011). The 40 face pairings were taken from an original set of 168 pairs (the full GFMT) and represent the most difficult trials (percentage correct $M = 81.3\%$, $SD = 9.7\%$; Burton et al., 2010).

The same forty pairs were recreated for the profile view version of the GFMT. These profile images (taken of the model's right side) were collected from the GFMT database materials (again, two different images of each face were available from this viewpoint), which the creators of the GFMT had additionally collected but were not used in the original face matching task. This allowed us to reproduce each face pair using the same identities and cameras as those used to collect the original frontal view images. For each frontal view trial, we created a profile version, keeping the same featured face(s), taken at the same camera distance with the same cameras, again with neutral expressions, showing the same clothing and hairstyles. The only difference was the view of the face itself.

All faces were shown in colour on a white background, and measured approximately 7 cm x 9 cm onscreen. Example stimuli are shown in Figure 1.



Figure 1. Example face pairs, where the top row shows a ‘match’ (the same woman; images taken with two different cameras) presented in frontal view, and the bottom row shows a ‘mismatch’ (two different men) presented in profile view.

Procedure

The 40 frontal view trials were presented together in one block, and the 40 profile view trials were presented together in a separate block. The order of these two blocks was counterbalanced across participants. The order of the trials within each block was randomised, as was the location of each face (left or right side) within each trial.

Because the frontal and profile versions of the GFMT contained the same identities and the same face-pairs, two steps were taken in order to minimise participants' attempts to remember/reproduce their responses. First, when completing the first block, participants were not told that there would be a second face-matching block later in the testing session. Second, participants completed a brief (5-15mins) distractor task between the two blocks. Although this task involved comparing face photos (card-sorting; Jenkins, White, Van Montfort, & Burton, 2011), there was no overlap between images or identities used in this task and the two face-matching blocks.

On each of the 40 trials in a block, two images were presented onscreen, one to the left and one to the right of centre, using custom MATLAB software. Viewing distance was not fixed. The task was to judge whether the two images were of the same person or two different people. Participants responded using the keyboard, pressing 'A' for 'same' and 'L' for 'different'. These labels remained onscreen throughout the experiment. Trials were self-paced, and no feedback was given at any point during the experiment.

Results

Comparing frontal and profile view performance

Overall percentage correct was analysed using a 2 x 2 mixed analysis of variance (ANOVA), where View (frontal, profile) varied within participants, and Block Order

(‘frontal then profile’, ‘profile then frontal’) varied between participants. Neither the main effect of View, $F(1, 52) = 1.54, p = .220, \eta^2_p = .03$, nor the main effect of Block Order, $F(1, 52) = 1.07, p = .307, \eta^2_p = .02$, were statistically significant. The interaction was also not significant, $F(1, 52) = 1.32, p = .256, \eta^2_p = .03$. Therefore, performance on frontal ($M = 84.95\%$, 95% CI [82.86%, 87.04%]) and profile views ($M = 83.70\%$, [81.43%, 85.98%]) did not differ in terms of percentage correct.

As with other research in this field (e.g., Kramer & Ritchie, 2016), we investigated signal detection measures in order to take into account potentially separate effects caused by the two views on ‘match’ and ‘mismatch’ trial types. We calculated sensitivity indices (d') using the following: *Hit*, both images are of the same identity and participants responded ‘same’; and *False alarm*, the two images are of different people and participants responded ‘same’. Trials from the two views were considered separately.

These d' values were analysed using a 2 x 2 mixed ANOVA as above. Neither the main effect of View, $F(1, 52) = 1.99, p = .165, \eta^2_p = .04$, nor the main effect of Block Order, $F(1, 52) = 0.32, p = .573, \eta^2_p = .01$, were statistically significant. The interaction was also not significant, $F(1, 52) = 1.32, p = .257, \eta^2_p = .03$. Therefore, performance on frontal ($M = 2.34$, 95% CI [2.16, 2.52]) and profile views ($M = 2.23$, [2.04, 2.42]) did not differ when considered in terms of d' , mirroring the results with percentage correct.

Finally, similar analyses were carried out for participants' response biases (criterion values). Here, we found a significant View x Block Order interaction, $F(1, 52) = 7.80, p = .007, \eta^2_p = .13$. We therefore carried out separate analyses for each level of View. We found no difference when comparing criterion values for frontal views when these were presented in the first ($M = 0.05, 95\% \text{ CI } [-0.13, 0.24]$) versus the second block ($M = -0.02, [-0.20, 0.17]$), $t(52) = 0.54, p = .592$, Cohen's $d = 0.15$. For profile views, we found a significant difference when comparing criterion values when these were presented in the first ($M = 0.25, [0.10, 0.41]$) versus the second block ($M = 0.02, [-0.13, 0.18]$), $t(52) = 2.06, p = .045$, Cohen's $d = 0.56$. However, we note that these response biases were relatively small, within normal ranges (Bobak, Dowsett, & Bate, 2016; Bobak, Hancock, & Bate, 2016), and were not investigated further here.

The observation of similar face matching performance (accuracy and sensitivity) for frontal and profile views suggests that both views contain similar amounts of information regarding identity. It also suggests that similar processes are involved when matching frontal and profile views. In order to test these conclusions, we ran two sets of follow-up comparisons.

Individual differences in performance across views

First, it is well established (using frontal views) that face matching ability appears to be a stable trait that generalises across tasks (e.g., Bobak, Dowsett, et al., 2016; Robertson,

Noyes, Dowsett, Jenkins, & Burton, 2016). For instance, Bobak Dowsett, et al. (2016) reported a correlation of .72 between participants' d' values on the frontal GFMT and a second, frontal face matching task. Therefore, if it is the case that similar processes are used when matching frontal views and profile views, then participants who performed well with frontal view face matching should also perform well with profile views. To test this prediction, we correlated performance across the two views for both percentage correct and d' sensitivity. Consistent with similar processes being used to match both views, we found large correlations for both measures: percentage correct, $r(52) = .57$, 95% CI [.36, .73], $p < .001$, and d' , $r(52) = .61$, [.41, .76], $p < .001$. (For additional analyses addressing the possibility of carryover effects across blocks, see the Supplementary Material.) Thus, the present data are consistent with similar processes being used when matching frontal and profile views. However, there is still substantial unexplained variability, suggesting that different information may be provided by the two views during face matching.

Face pair difficulty across views

The second follow-up analysis assessed whether face pairs that were difficult when presented in frontal view were also difficult when viewed in profile. Again, if frontal and profile views are processed similarly, using similar information, then performance should be similar for the two views. Because the same face pairs were presented in both

the frontal view and profile view conditions, it was possible to test this prediction by correlating the accuracy (percentage correct) for face pairs in frontal view with their profile view equivalents (collapsing across participants) for the 40 pairs. Consistent with similar processes being used for both views, there was a large correlation, $r(38) = .49$, 95% CI [.21, .69], $p = .001$, demonstrating that difficulty for face pairs presented in the two views is related. However, as above, there is still substantial unexplained variability, suggesting that the two views may provide different information during face matching. (For additional analyses addressing the possibility of carryover effects across blocks, see the Supplementary Material.)

Experiment 2

Is face matching improved when both frontal and profile views are used simultaneously, compared with when a single view is presented? The results of Experiment 1 suggest that presenting both views should improve performance. Although face matching performance was similar for frontal and profile views, the correlational analyses suggested that different information may have been used when matching faces. Data from ‘across views’ face matching (comparing a frontal image to a profile image) studies support this conclusion. Face matching across views leads to a decrease in performance (approximately 3-10%) in comparison with frontal-frontal image pairs (Bindemann, Attard, Leach, & Johnston, 2013; Estudillo & Bindemann, 2014). Again,

this suggests that different information is available in the two views. However, there is some evidence to suggest that face recognition (which involves a memory component) is not improved when both frontal and profile views are shown together at learning and then one of these views is shown at test (Logie et al., 1987). This makes it unclear whether people will take advantage of the fact that the two views provide different information.

In order to assess whether presenting both views simultaneously increases matching performance in comparison with each view presented individually, participants in Experiment 2 performed face matching using the same face pairs as in Experiment 1. Unlike Experiment 1, however, two views (frontal and profile) were presented for each face. Performance was then compared to the Block 1 frontal and profile view performance from Experiment 1, creating a between subjects design. If face matching performance is better in the ‘two views’ condition, this would suggest that there is different information in each view *and* that participants can combine these sources to facilitate matching.

Methods

Participants

Thirty-three students (31 women; age $M = 18.76$ years, $SD = 1.77$; 87.88% self-reported ethnicity as White) at Trent University took part in exchange for course credits. None of

these participants had taken part in Experiment 1. Sample size was based on the number of participants who saw each view in their first block in Experiment 1 (i.e., 27), although we continued to test participants until the end of a pre-determined data collection run.

Stimuli

The same stimuli were used here as in Experiment 1. Again, all faces were shown in colour on a white background, and measured approximately 7 cm x 9 cm onscreen. An example trial is shown in Figure 2.

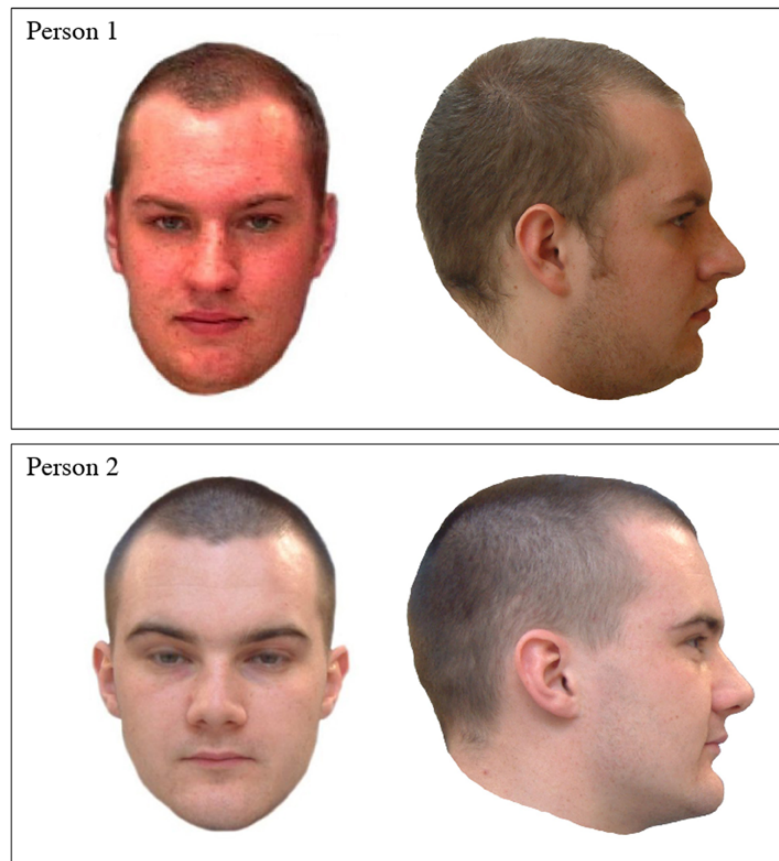


Figure 2. An example trial from Experiment 2, where each box shows a face in both frontal and profile views. Participants are asked to compare the two boxes and decide whether they show the same person or two different people. Here, two different men are depicted.

Procedure

The procedure was similar to Experiment 1, except that participants only saw one block of 40 trials. The order of the trials within this block was randomised.

On each of the 40 trials, four face images were presented onscreen, one in each quadrant, using custom MATLAB software. These images comprised a frontal and a profile view for each face. Boxes and text appeared onscreen (see Figure 2) to illustrate how the four images represented two image pairs. The task was to judge whether the two pairs of images were of the same person or two different people. Participants were shown an example of the trial layout on paper beforehand (a ‘mismatch’ trial, showing frontal and profile views of Tom Cruise and Ryan Gosling) in order to make certain that the instructions were clear, and so that the experimenter could answer any questions that might arise. Participants responded using the keyboard, pressing ‘A’ for ‘same’ and ‘L’ for ‘different’. These labels remained onscreen throughout the experiment. Trials were self-paced, and no feedback was given at any point during the experiment. Viewing distance was not fixed.

Results

Comparing performance with frontal, profile, and both views

Data from Block 1 of Experiment 1 were combined with the data collected here. As such, 27 participants saw only frontal view trials, 27 saw only profile view trials, and 33 saw both views simultaneously (with no overlap across these three participant samples).

Overall percentage correct was analysed using a univariate ANOVA, where View (frontal, profile, both) varied between participants. The main effect of View was not

statistically significant, $F(2, 84) = 2.55, p = .084, \eta^2_p = .06$. We also found no main effect of View when we analysed d' values, $F(2, 84) = 2.13, p = .126, \eta^2_p = .05$, and criterion values, $F(2, 84) = 1.85, p = .163, \eta^2_p = .04$. A summary of these results can be seen in Table 1.

View	<i>n</i>	Percentage Correct	d'	Criterion
Frontal	27	85.37 [82.18, 88.56]	2.34 [2.07, 2.62]	0.05 [-0.10, 0.21]
Profile	27	82.13 [78.94, 85.32]	2.14 [1.86, 2.41]	0.25 [0.10, 0.41]
Both	33	86.97 [84.09, 89.85]	2.52 [2.27, 2.77]	0.21 [0.07, 0.34]

Table 1. A summary of the data for the three viewing conditions in Experiment 2.

Values represent the means, with 95% confidence intervals shown in square brackets.

Face pair difficulty across views

If similar information is used to match faces in each of the three conditions then face pair accuracies (percentage correct) should be similar across the three viewing conditions. As in Experiment 1, we tested this prediction by calculating an accuracy score for each of the face pairs in each view by averaging across participants and then correlating the accuracies for the three views (the same 40 face pairs in each view). We found significant correlations in all cases: frontal–profile, $r(38) = .39$, 95% CI [.09, .62], $p = .013$; frontal–both, $r(38) = .60$, [.35, .77], $p < .001$; profile–both, $r(38) = .64$, [.42,

.80], $p < .001$. Interestingly, these three correlations did not differ statistically (all $ps > .136$). Again, the observation of significant correlations suggests that at least some of the same visual features or properties are used in each view when comparing the faces. The observation that the correlations did not differ between frontal-both and profile-both is also consistent with similar image features being used in both conditions.

Are both views used when simultaneously displayed?

Although the correlations are suggestive, they are far from conclusive. Therefore, two regression analyses were conducted on the face-pair accuracies in order to gain more insight into how participants used information from the frontal and profile views (see Table 2). Our Theoretical Model considered the ‘frontal only’ and ‘profile only’ conditions, as well as their interaction, as predictors, and the ‘both views’ condition as the outcome variable. The overall model was significant, $F(3, 36) = 19.14, p < .001$, with an R^2 of .62. Furthermore, consistent with participants using information from both images when making decisions, both frontal accuracy and profile accuracy explained some unique variance in the ‘both views’ accuracy, indicating that both views were being used during the experiment. Critically, the interaction was also significant, suggesting that the information from both views was being considered on each trial. When accuracy in the frontal condition was high (1 *SD* above average), the slope relating profile accuracy and the ‘both views’ accuracy was not significant, $\beta = .18, p =$

.202, and when frontal accuracy was low ($-1\ SD$), this slope was significant, $\beta = .60$, $p < .001$. This result suggests that profile information may have been utilised more when frontal comparisons were difficult.

Predictor	<i>B</i>	β	<i>R</i> ²
Theoretical Model			.615
Intercept	0.88 [0.85, 0.91]**		
Frontal accuracy	0.38 [0.11, 0.66]**	0.33 [0.10, 0.57]	
Profile accuracy	0.39 [0.19, 0.60]**	0.45 [0.22, 0.68]	
Frontal x profile	-2.05 [-3.81, -0.28]*	-0.26 [-0.49, -0.04]	
Control Model			.507
Intercept	-0.01 [-0.29, 0.28]		
Highest of accuracies	0.99 [0.67, 1.31]**	0.71 [0.48, 0.94]	

Table 2. Regression models where ‘both views’ accuracy was the outcome variable.

Both unstandardised (*B*) and standardised (β) coefficients are presented, with square brackets providing 95% confidence intervals. * indicates $p < .05$; ** $p < .01$.

Before drawing conclusions, we wanted to compare this model to a reasonable control model. We therefore considered the case where, rather than integrating the information from both the frontal and profile views, participants relied on a single view

for each trial (even though both were presented), and this view changed, depending on the trial itself. Here, we examined a model where the participant somehow knew to use the view that yielded the greatest accuracy on its own on each trial. For example, if the accuracy for a particular face pair was 80% for the frontal view and 70% for the profile view, then only the frontal view accuracy would be considered. To test this ‘highest of two accuracies’ Control Model, we ran a second regression analysis. For this analysis, the predictor was a new variable calculated as the highest of the accuracies found for each of the two single views. Again, accuracy in the ‘both views’ condition was the outcome variable. The model was significant, $F(1, 38) = 39.04, p < .001$, explaining 51% of the variance in how participants actually performed when both views appeared together. Critically, the Control Model had less predictive power than the Theoretical Model, which explained 62% of the variance in ‘both views’ accuracy, $F(2, 36) = 5.04, p = .012$.

The conclusion that participants were integrating information from the frontal and profile views on each trial is consistent with them performing optimally. It is therefore surprising that performance in the ‘both views’ condition was no higher than single view conditions. The data from Experiment 1 indicated that the accuracy for frontal and profile views only shared approximately 15.1% of their variance (the correlation, reported above, was .39). This suggests that there should be a substantial amount of unique information that each could contribute.

Experiment 3

The results of Experiment 2 show that accuracy and d' across the three viewing conditions did not differ, and hence providing frontal and profile views simultaneously failed to produce an improvement in performance. However, Experiment 2 incorporated data from the first experiment in order to compare across the three viewing conditions, and as such, did not provide a standalone comparison of performance across the three conditions. Further, although no effect of View was found for either overall percentage correct ($p = .084$) or d' values ($p = .126$), at least the former result might cause some readers to question whether a difference may still be present.

In addition, our regression analyses of Experiment 2 suggested that participants did seem to integrate the information from both views, and yet performance was no higher in the 'both views' condition. Given this perhaps counterintuitive result, we sought to replicate the experiment with a larger sample size and hence an increased power to detect such effects. To foreshadow, the larger sample size used here again supported the idea that both views were considered by participants but failed to result in an improvement in performance.

In Experiment 3, we aimed to address these issues with a full conceptual replication of Experiment 2. Here, new data for all three views were collected (again between subjects), using sample sizes that were approximately twice those of

Experiment 2. Further, we collected response time data in order to provide additional insight into how participants viewed the images. To anticipate the results, we again find no difference in performance across the three viewing conditions.

Methods

Participants

A community sample of 167 participants (77 women; age $M = 35.13$ years, $SD = 10.77$; 70.06% self-reported ethnicity as White) were obtained via Amazon's Mechanical Turk in exchange for \$0.50 in payment. None of these participants had taken part in the previous experiments. All participants provided informed consent online and were shown a debriefing screen at the end of the experiment. Sample size was chosen in order to be approximately twice the size of the groups analysed in Experiment 2.

The data from 12 additional participants were excluded before analyses for reasons explained below. However, all participants were unique (due to the nature of the project specifications on MTurk) and so no data were excluded because of repeated participation by the same individuals.

Stimuli

The same stimuli were used here as in Experiments 1 and 2. Again, all faces were shown in colour on a white background. The absolute size of the images onscreen was

kept approximately the same for all participants through the use of a calibration procedure, which took place before the experiment, scaling the stimuli to the user's specific screen resolution. In addition, due to the experimental software used, participants were prevented from taking part on mobile devices (i.e., very small screens).

Procedure

The procedure was similar to Experiment 2 in that participants only saw one block of 40 trials. Each participant was randomly assigned to a viewing condition: frontal, profile, or both views. The order of the trials was randomised.

The experiment was completed online through the Testable website (www.testable.org). First, participants were instructed to set their browsing windows to full screen, minimise possible distractions (e.g., TV, phone, etc.), and position themselves at arm's length from the monitor for the duration of the experiment (although viewing distance was not fixed). Next, a screen size calibration took place (adjusting an onscreen bar to match the length of a credit card), consent was obtained, and then demographic information was collected.

As with Experiments 1 and 2, onscreen instructions were provided. In addition, for the 'both views' condition, participants were taken through two examples (one 'same' and one 'different' trial) using celebrities' faces in order to make sure that the

instructions were understood. Two examples were provided here (vs. one in Experiment 2) in order to make up for the fact that online participants were unable to ask clarifying questions of the experimenter. Participants responded using the keyboard, pressing ‘A’ for ‘same’ and ‘L’ for ‘different’. These labels remained onscreen throughout the experiment. Trials were self-paced, and no feedback was given at any point during the experiment. In addition, the Testable platform recorded how long participants took to respond on each trial.

In order to check whether participants were concentrating on the task (since this can be a concern for online studies), we included one additional trial that always appeared at the end of the experiment. One image from the first trial (or two images of an identity in the ‘both’ condition) was presented again, but this time paired with itself. Because these were two identical images (or pairs of images), participants were expected to respond ‘same’. The images in this trial were presented in the same view as the other trials for that participant.

Results

Data from 10 participants were excluded because they responded incorrectly to the final, image-matching trial. Furthermore, in the original GFMT study (Burton et al., 2010), none of the 194 participants scored lower than 51%. We therefore excluded two additional participants because their overall accuracies (percentage correct) were below

50% (chance). In total, 12 participants were excluded: frontal view (3), profile view (4), and both views (5).

Comparing performance with frontal, profile, and both views

Percentage correct was analysed using a univariate ANOVA, where View (frontal, profile, both) varied between participants. The main effect of View was not statistically significant, $F(2, 164) = 2.50, p = .085, \eta^2_p = .03$. We also found no main effect of View when we analysed d' values, $F(2, 164) = 1.90, p = .153, \eta^2_p = .02$, and criterion values, $F(2, 164) = 0.04, p = .965, \eta^2_p = .00$. A summary of these results can be seen in Table 3. It should be noted that here, the frontal view yielded the lowest average score, whereas in Experiment 2 it was the profile view.

View	<i>n</i>	Percentage Correct	d'	Criterion
Frontal	51	76.62 [73.34, 79.90]	1.75 [1.50, 1.99]	0.07 [-0.07, 0.20]
Profile	57	81.10 [77.99, 84.20]	2.02 [1.79, 2.25]	0.07 [-0.06, 0.19]
Both	59	81.06 [78.01, 84.11]	2.05 [1.82, 2.27]	0.05 [-0.08, 0.17]

Table 3. A summary of the data for the three viewing conditions in Experiment 3.

Values represent the means, with 95% confidence intervals shown in square brackets.

Face pair difficulty across views

As with Experiments 1 and 2, we tested whether face pair accuracies (percentage correct) were similar across the three viewing conditions by calculating an accuracy score for each of the face pairs in each view by averaging across participants, and then correlating the accuracies for the three views (the same 40 face pairs in each view). We found significant correlations in all cases: frontal–profile, $r(38) = .42$, 95% CI [.12, .65], $p = .007$; frontal–both, $r(38) = .76$, [.59, .87], $p < .001$; profile–both, $r(38) = .55$, [.29, .74], $p < .001$. As with Experiment 2, the observation of significant correlations suggests that at least some of the same information is used in each view. Again, the observation that the correlations did not differ between frontal–both and profile–both ($p = .110$) is also consistent with similar information being used in both conditions. While we find no difference between the frontal–profile correlation and the profile–both correlation ($p = .447$), we do find that the frontal–both correlation is significantly larger than the frontal–profile correlation ($p = .019$).

Are both views used when simultaneously displayed?

We repeated the regression analyses performed in Experiment 2 (see Table 4). In our Theoretical Model, the ‘frontal only’ and ‘profile only’ conditions and their interaction were used as predictors, and the ‘both views’ condition was used as the outcome variable. If the frontal and profile views are both contributing to performance in the ‘both views’ condition, then each should explain some unique variance in the ‘both

views' accuracy. Furthermore, if the information in both views is being used on each trial, then the interaction should also account for some unique variance. The overall model was significant, $F(3, 36) = 24.89, p < .001$, with an R^2 of .68. Consistent with participants using information from the frontal images when making decisions, frontal accuracy was a significant predictor. However, neither profile accuracy nor the interaction were significant predictors. Indeed, the addition of the interaction term provided no additional explanatory power above a model including only the two main effects, with a nonsignificant R^2 change, $F(1, 36) = 3.53, p = .069$. It is worth noting that by excluding the interaction term, 'profile view' accuracy became a significant predictor ($\beta = .284, p = .012$), and resulted in a significant improvement over a 'frontal view' only model, $F(1, 37) = 6.91, p = .012$. That said, we discuss the interaction to maintain consistency with Experiment 2. Once again, when frontal accuracy was high (+1 SD), there was no effect of profile accuracy, $\beta = .078, p = .573$, but when frontal accuracy was low (-1 SD) there was a significant effect of profile accuracy, $\beta = .310, p = .004$.

Predictor	<i>B</i>	β	R^2
Theoretical Model			.675
Intercept	0.82 [0.80, 0.83]**		
Frontal accuracy	0.51 [0.32, 0.70]**	0.58 [0.36, 0.80]	

Profile accuracy	0.19 [-0.02, 0.40]	0.21 [-0.02, 0.44]	
Frontal x profile	-1.09 [-2.27, 0.09]	-0.21 [-0.44, 0.02]	
Control Model			.478
Intercept	0.24 [0.05, 0.44]*		
Highest of accuracies	0.68 [0.45, 0.92]**	0.69 [0.45, 0.93]	
RT Theoretical Model			.542
Intercept	0.82 [0.80, 0.85]**		
Frontal RT	-0.20 [-0.32, -0.08]**	-0.43 [-0.68, -0.18]	
Profile RT	-0.11 [-0.20, -0.02]*	-0.31 [-0.56, -0.06]	
Frontal RT x profile RT	-0.67 [-1.05, -0.29]**	-0.40 [-0.63, -0.17]	

Table 4. Regression models where ‘both views’ accuracy was the outcome variable.

Both unstandardised (B) and standardised (β) coefficients are presented, with square brackets providing 95% confidence intervals. * indicates $p < .05$; ** $p < .01$.

In the Control Model (introduced in Experiment 2), the predictor we used was a new variable calculated where the predicted accuracy for each trial was based on the single view with the highest accuracy. This model represented an imagined participant who only used the most informative view (frontal or profile) on each trial. Again, accuracy in the ‘both views’ condition was the outcome variable. The Control Model was significant, $F(1, 38) = 34.76$, $p < .001$, explaining 48% of the variance in how

participants actually performed when both views appeared together. Again, the Control Model had less predictive power than the Theoretical Model, which explained 68% of the variance in the ‘both views’ condition, $F(2, 36) = 10.90, p < .001$. This result suggests that participants were using information from both views rather than simply considering the most informative one. Evidence that participants were using both views in their decisions begs the question: if both views are considered, and they provide non-overlapping information, then why do we find no performance increase?

If participants used both views where available then we would predict that response times would be longer for this condition in comparison with each individual view. The Testable website recorded how long participants spent on each face pair, and so we calculated the average time spent (across pairs) for each participant. These average response times were analysed using a univariate ANOVA, where View (frontal, profile, both) varied between participants. The main effect of View was statistically significant, $F(2, 164) = 4.45, p = .013, \eta^2_p = .05$. Follow-up comparisons (Bonferroni corrected) revealed that response times for frontal ($M = 2.77s$) and profile views ($M = 2.88s$) did not differ ($p = 1.00$), while both these views resulted in shorter response times than in the ‘both views’ condition ($M = 4.03s$; both $ps < .043$). That participants spent longer when presented with both views suggests that all images were considered, supporting the above finding that both views contribute to accuracy rather than simply the most informative single view. It is worth noting, however, that these results may be

specific to situations in which viewing time is unlimited, and that decisions made under time constraints could see participants restricting their comparisons and failing to consider all the images.

In order to better understand the RT data, we examined how it predicted trial accuracy using the Theoretical Model (see RT Theoretical Model in Table 4). Accuracy in the ‘both views’ condition was the outcome variable, with single view response times for the frontal trials, profile trials, and their interaction, as the predictor variables. Given that RT and accuracy are typically inversely related (easy tasks have high accuracy and short response times), we predicted that RT would be negatively related to accuracy. The overall model was significant, $F(3, 36) = 14.19, p < .001$, with an R^2 of .54. As can be seen in the table, all predictors yielded significant effects. As RT increased, accuracy decreased. Further analysis of the interaction revealed that when frontal RTs were fast, there was no effect of profile RTs ($B = .023, p = .695$), but when frontal RTs were slow, there was a significant effect of profile RTs ($B = -.245, p < .001$). This is consistent with participants using the profile view only when the frontal view did not yield a definitive answer.

No evidence of an other-race effect

Given the mixture of ethnicities and the relatively large sample size here, we considered the possibility that non-White participants may have shown worse face matching

performance since the stimuli depicted White individuals (Megreya, White, & Burton, 2011), and that this, in turn, could have masked any potential effects of View. We therefore analysed the d' values using a two-way ANOVA, where both View (frontal, profile, both) and Ethnicity (White, non-White) varied between participants. The main effect of View was not statistically significant, $F(2, 161) = 1.02, p = .362, \eta^2_p = .01$. We also found no main effect of Ethnicity, $F(1, 161) = 0.39, p = .532, \eta^2_p = .00$, and no interaction between the two factors, $F(2, 161) = 0.15, p = .862, \eta^2_p = .00$. Therefore, at least in the current work, and with a non-White sample that was limited in size, we found no evidence of an other-race detriment in face matching.

General Discussion

The present experiments addressed two questions. The first question was “Do frontal views lead to more accurate face matching than profile views?” The results of Experiment 1 showed that levels of performance did not differ for the frontal and profile views, suggesting that profile-profile and frontal-frontal image matching were similarly difficult. The results of Experiment 3 replicated this finding. The second question was “Can information from frontal and profile views be combined to produce more accurate face matching?” Experiments 2 and 3 showed that performance was not improved by presenting both views simultaneously. Follow-up regression analyses of the data from Experiments 2 and 3 suggested that performance was not improved when both views

were presented simultaneously, despite evidence that participants were integrating information in each view and these views provided at least some unique information.

The observation that face matching performance did not differ for frontal and profile views seems surprising, given the disproportionate emphasis given to frontal views in identification documents (though there is widespread collection of profile views in standardised mugshots). That said however, there is some evidence that frontal and profile views can be equally informative in other domains. For instance, previous research suggests a large overlap in attractiveness signals from frontal and profile views of the same face (Valenzano, Mennucci, Tartarelli, & Cellerino, 2006). In addition, facial expressions are recognised equally well from both views (Matsumoto & Hwang, 2011).

Here, we found some evidence that similar processes were used when face matching using the frontal views and profile views. For instance, in Experiment 1, we found a correlation between participant performances on each view. That is, people who are good at face matching with frontal images also tend to be good with profile images. This suggests an underlying ability that varies across, but is stable within, individuals. Previous research supports this idea, where those who perform well or poorly tend to do so across different face tasks (Bobak, Dowsett, et al., 2016; Robertson et al., 2016; Verhallen et al., 2017).

Despite evidence that the frontal and profile views shared similar processes, we found that a large proportion of variance in face pair accuracy is not explained by the overlap between views (the percentage correct correlation between the two views was .49 in Experiment 1), suggesting that frontal and profile views provide unique identity information. If participants are comparing simple visual features then each view may provide different information regarding these features. For instance, the width of a nose is easily discerned in a frontal view, but not in a profile view, whereas the opposite pattern is found for the length of a nose. Other information, such as a pronounced brow ridge would be visible in both views. However, investigating and quantifying this overlap requires experiments specifically tailored to this question, and represents an interesting path for further research.

Although there was evidence that the frontal and profile views provided different information, we observed no evidence that this information could be used to improve performance in the ‘both views’ condition. To our knowledge, there is no work to date investigating how people compare two pairs of images. However, there is some evidence to suggest that face recognition (which involves a memory component) is not improved when both frontal and profile views are shown together at learning (vs. only one of these views) and then one of these views is shown at test (Logie et al., 1987). For face matching, there is evidence suggesting that comparing a single image to an array of

unconstrained images improves performance, although only on ‘match’ trials (White et al., 2014).

If the frontal and profile views of the face provide distinct and useful information for matching as the present data suggest, then the failure to observe better performance when both views are presented provides a challenge for any theory of face matching. For instance, previous research in ensemble encoding has shown that several instances of a face are automatically averaged to form what is thought to be a more useful representation of that individual (Jenkins & Burton, 2008; Kramer, Ritchie, & Burton, 2015). Although simple averaging of the frontal and profile images of a face presented here would have resulted in a noisy and crude image, some level of integration may have still been possible. For instance, the two views could be combined to form the beginnings of a three-dimensional representation of the face. However, if this were the case then some improvement in performance would have been expected. That Experiments 2 and 3 found no improvement when both views were presented simultaneously leaves us with no evidence that any form of integration took place here.

A logical follow-up, given the above research, may be to somehow encourage participants to construct an internal representation, combining information from both views for each face, since this may lead to improvements in matching performance. For example, presenting both views for a given face and then removing these from the screen might force viewers to construct a representation of the face for use in

subsequent comparisons. Future research should investigate how and when participants shift their comparisons from image-based to representation-based.

In the current work, rather than forming an internal representation of the face, it is perhaps more likely that participants' comparisons were driven by simple image feature matching, where the properties of the images represent the only available information for unfamiliar faces (Hancock et al., 2000). Rather than forming an internal, view-independent representation of the face for each identity, it may have been simpler (with so few images) to separately compare the two frontal images and the two profiles, without any attempt to integrate the views. Although no view-independent representation is formed, information from both view comparisons can still be integrated (e.g., the frontal views may suggest two people with different noses, and so this forms a working hypothesis when consulting the profiles). Consistent with this account, we found some evidence that each of the views contributed to performance in the 'both views' condition (the regression in Experiment 3). In order to explore this idea further, a future study might, for example, consider asking participants to respond based upon each view separately, and then to provide a final decision when presented with both views. Eye-tracking techniques may also shed some light on how participants approach such comparisons.

For recognition memory across views, the important factor appears to be the angular difference between learning and test views (Baddeley & Woodhead, 1983; Liu

& Chaudhuri, 2002). The present results suggest that a similar effect may be occurring in face matching. For face matching, we know that comparing a frontal to a profile image leads to a decrease in performance (also using the GFMT; Estudillo & Bindemann, 2014), while keeping the view unchanged produces the same accuracy no matter the view (shown here). However, researchers have yet to investigate whether angular difference plays the same crucial role as in recognition. By collecting frontal, three-quarter, and profile views of each face in future, and then investigating all combinations of these during face matching, researchers might find support for this idea.

In conclusion, we show that unfamiliar face matching performance does not differ when both faces are presented frontally versus in profile. In addition, we find no benefit when participants are presented with both frontal and profile images for each face simultaneously. These results raise two interesting questions: do frontal and profile views provide different information for identification, and if they do, why are participants unable to utilise both sources? Our initial motivation for the current work was the possibility that profile images might be used for identification, either instead of or alongside frontal images. However, our results categorically discount both changes, with neither producing an improvement over traditional, frontal image matching.

Acknowledgements

The authors thank Mike Burton for his advice during the earlier stages of this project.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

References

- Baddeley, A., & Woodhead, M. (1983). Improving face recognition ability. In S. M. A. Lloyd-Bostock & B. R. Clifford (Eds.), *Evaluating witness evidence* (pp. 125-136). Chichester: Wiley.
- Bindemann, M., Attard, J., Leach, A., & Johnston, R. A. (2013). The effect of image pixelation on unfamiliar-face matching. *Applied Cognitive Psychology*, 27(6), 707-717.
- Bobak, A. K., Dowsett, A. J., & Bate, S. (2016). Solving the border control problem: Evidence of enhanced face matching in individuals with extraordinary face recognition skills. *PLoS ONE*, 11(2), e0148148.

- Bobak, A. K., Hancock, P. J. B., & Bate, S. (2016). Super-recognisers in action: Evidence from face-matching and face memory tasks. *Applied Cognitive Psychology*, 30(1), 81-91.
- Bruce, V., Henderson, Z., Greenwood, K., Hancock, P. J. B., Burton, A. M., & Miller, P. (1999). Verification of face identities from images captured on video. *Journal of Experimental Psychology: Applied*, 5(4), 339-360.
- Bruce, V., Henderson, Z., Newman, C., & Burton, A. M. (2001). Matching identities of familiar and unfamiliar faces caught on CCTV images. *Journal of Experimental Psychology: Applied*, 7(3), 207-218.
- Burton, A. M., White, D., & McNeill, A. (2010). The Glasgow Face Matching Test. *Behavior Research Methods*, 42, 286-291.
- Dowsett, A. J., & Burton, A. M. (2015). Unfamiliar face matching: Pairs out-perform individuals and provide a route to training. *British Journal of Psychology*, 106, 433-445.
- Estudillo, A. J., & Bindemann, M. (2014). Generalization across view in face memory and face matching. *i-Perception*, 5, 589-601.
- Hancock, P. J. B., Bruce, V., & Burton, A. M. (2000). Recognition of unfamiliar faces. *Trends in Cognitive Sciences*, 4(9), 330-337.
- Jenkins, R., & Burton, A. M. (2008). 100% accuracy in automatic face recognition. *Science*, 319(5862), 435.

- Jenkins, R., & Burton, A. M. (2011). Stable face representations. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366, 1671–1683.
- Jenkins, R., White, D., Van Montfort, X., & Burton, A. M. (2011). Variability in photos of the same face. *Cognition*, 121(3), 313–323.
- Kemp, R. I., Caon, A., Howard, M., & Brooks, K. R. (2016). Improving unfamiliar face matching by masking the external facial features. *Applied Cognitive Psychology*, 30(4), 622–627.
- Kramer, R. S. S., & Ritchie, K. L. (2016). Disguising Superman: How glasses affect unfamiliar face matching. *Applied Cognitive Psychology*, 30(6), 841–845.
- Kramer, R. S. S., Ritchie, K. L., & Burton, A. M. (2015). Viewers extract the mean from images of the same person: A route to face learning. *Journal of Vision*, 15(4):1, 1–9.
- Liu, C. H., & Chaudhuri, A. (2002). Reassessing the 3/4 view effect in face recognition. *Cognition*, 83(1), 31–48.
- Logie, R. H., Baddeley, A. D., & Woodhead, M. M. (1987). Face recognition, pose and ecological validity. *Applied Cognitive Psychology*, 1, 53–69.
- Matsumoto, D., & Hwang, H. S. (2011). Judgments of facial expressions of emotion in profile. *Emotion*, 11(5), 1223–1229.
- Megreya, A. M., & Burton, A. M. (2006). Unfamiliar faces are not faces: Evidence from a matching task. *Memory & Cognition*, 34(4), 865–876.

- Megreya, A. M., & Burton, A. M. (2008). Matching faces to photographs: Poor performance in eyewitness memory (without the memory). *Journal of Experimental Psychology: Applied*, 14(4), 364-372.
- Megreya, A. M., White, D., & Burton, A. M. (2011). The other-race effect does not rely on memory: Evidence from a matching task. *The Quarterly Journal of Experimental Psychology*, 64(8), 1473-1483.
- Robertson, D. J., Noyes, E., Dowsett, A. J., Jenkins, R., & Burton, A. M. (2016). Face recognition by Metropolitan Police super-recognisers. *PLoS ONE*, 11(2): e0150036.
- Valenzano, D. R., Mennucci, A., Tartarelli, G., & Cellerino, A. (2006). Shape analysis of female facial attractiveness. *Vision Research*, 46(8), 1282-1291.
- Verhallen, R. J., Bosten, J. M., Goodbourn, P. T., Lawrance-Owen, A. J., Bargary, G., & Mollon, J. D. (2017). General and specific factors in the processing of faces. *Vision Research*. Advance online publication.
- White, D., Burton, A. M., Jenkins, R., & Kemp, R. I. (2014). Redesigning photo-ID to improve unfamiliar face matching performance. *Journal of Experimental Psychology: Applied*, 20(2), 166-173.
- White, D., Burton, A. M., Kemp, R. I., & Jenkins, R. (2013). Crowd effects in unfamiliar face matching. *Applied Cognitive Psychology*, 27(6), 769-777.